Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins

Benjamin Schuler, Andrea Soranno, Hagen Hofmann, and Daniel Nettels

Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland; email: schuler@bioc.uzh.ch

Abstract

The properties of unfolded proteins have long been of interest because of their importance to the protein folding process. Recently, the surprising prevalence of unstructured regions or entirely disordered proteins under physiological conditions has led to the realization that such intrinsically disordered proteins can be functional even in the absence of a folded structure. However, owing to their broad conformational distributions, many of the properties of unstructured proteins are difficult to describe with the established concepts of structural biology. We have thus seen a reemergence of polymer physics as a versatile framework for understanding their structure and dynamics. An important driving force for these developments has been single-molecule spectroscopy, as it allows structural heterogeneity, intramolecular distance distributions, and dynamics to be quantified over a wide range of timescales and solution conditions. Polymer concepts provide an important basis for relating the physical properties of unstructured proteins to folding and function.

Keywords

single-molecule fluorescence, correlation spectroscopy, protein dynamics, protein folding, internal friction
INTRODUCTION

Polymer physics has a long history in biology and was applied to biopolymers long before the routine availability of high-resolution structural information from crystallography or nuclear magnetic resonance (NMR). For proteins, however, the compelling atomistic detail provided by structural biology resulted in a predominant interest in their precise three-dimensional structures. Only in some areas (e.g., protein folding) did polymer ideas continue to be employed and developed (20, 45, 73, 141). A recent change in trend has been stimulated by the increasing awareness that unstructured regions or entirely disordered proteins are surprisingly widespread, especially in eukaryotes (164). For these intrinsically disordered proteins (IDPs), polymer properties—such as structural heterogeneity, broad conformational distributions, and long-range fluctuations—are of obvious relevance (5).

A broad range of powerful methods for investigating unfolded proteins and IDPs exist (162), among them NMR (75) and small-angle X-ray scattering (SAXS) (9). Recently, single-molecule spectroscopy has developed into a complementary approach for probing such systems (15, 22, 46) for several reasons: It has the capability of resolving structural and dynamic heterogeneity and avoiding many complications of ensemble averaging (46); it enables access to a wide spectrum of timescales for investigating dynamics (137); and it allows investigations over a broad range of solution conditions, from very low sample concentrations, where nonidealities such as aggregation are negligible, to highly heterogeneous or crowded environments, including live cells (1, 83, 130). The two predominant types of single-molecule techniques used for biomolecules and suitable for probing their polymer properties are force (126, 182) and fluorescence spectroscopy. Here we focus on work that uses fluorescence spectroscopy, particularly in combination with Förster resonance energy transfer (FRET).

SINGLE-MOLECULE FRET OF UNFOLDED AND DISORDERED PROTEINS

Single-molecule FRET is an attractive method for probing long-range distances and distance dynamics in unfolded proteins because it is most sensitive in the range of ~2 to 10 nm, an
intramolecular distance range typical of polypeptide segments from several tens of to a few hundred residues, depending on the compactness of the chain. The theoretical basis of the FRET process was established by Theodor Förster (51), who showed that the rate coefficient of excitation energy transfer, \( k_T \), between a suitable donor and acceptor chromophore is proportional to the inverse sixth power of their distance,

\[
k_T = k_D \left( \frac{R_0}{r} \right)^6,
\]

where \( k_D^{-1} = \tau_D \) is the fluorescence lifetime of the donor in the absence of the acceptor (Figure 1), and \( r \) is the distance between donor and acceptor. \( R_0 \), the Förster radius, can be obtained from readily measurable quantities (51): the overlap integral between the donor emission and the acceptor absorption spectra, the donor fluorescence quantum yield, the refractive index of the medium between the dyes, and the relative orientation of the chromophores. For cases in which the rotational reorientation of the chromophores is fast compared to the donor excited-state lifetime, the orientational contribution, \( \kappa^2 \), averages to a value of 2/3, which simplifies the application of FRET considerably [the applicability of \( \kappa^2 = 2/3 \) can be tested by polarization-sensitive detection (34, 146)]. Assuming a single fixed distance, the probability that a photon absorbed by the donor will lead to energy transfer to the acceptor, called the FRET efficiency, \( E \), is given by \( k_T/(k_T + k_D) \), which (with Equation 1) leads to the relation (Figure 1)

\[
E(r) = \frac{R_0^6}{R_0^6 + r^6}.
\]

The Förster radius is thus the characteristic distance that results in a transfer efficiency of 1/2. \( E \) is determined experimentally either by counting the number of emitted donor and acceptor photons \( [E = n_A/(n_A + n_D)] \), including the necessary corrections (34) or by measuring fluorescence lifetimes, for example, of the donor in the presence (\( \tau_{DA} \)) and absence (\( \tau_D \)) of the acceptor (\( E = 1 - \tau_{DA}/\tau_D \) for a single fixed distance).

Because the molecules are observed individually (either from bursts of photons originating from single molecules diffusing through the confocal observation volume, or in fluorescence time traces of immobilized molecules), structural heterogeneity can be resolved if the underlying conformations differ in transfer efficiency and if their interconversion dynamics are sufficiently slow compared to the time resolution of the measurement. A simple example is the equilibrium between folded and unfolded molecules freely diffusing in solution, resulting in separate subpopulations in the transfer efficiency histogram (Figures 1c and 3a) if they interconvert on a timescale slower than approximately a millisecond, the typical diffusion time through the confocal volume. In this case, the events corresponding to unfolded molecules can be singled out, and the properties of the unfolded state can be investigated without overlap from the signal of folded molecules (and of molecules lacking an active acceptor dye). The same idea can be extended to systems with more than two subpopulations (8, 121). For a more detailed introduction to the technical aspects and the versatility of multiparameter detection and analysis, we refer the reader to recent reviews (136 and 146).

**Quantifying Distance Distributions**

An important aspect that needs to be taken into account for unfolded proteins and IDPs is that we are not dealing with a single fixed intramolecular distance, but with a distribution of distances, \( P(r) \). Even if we have separated the fluorescence events of, for example, folded and unfolded molecules, we still have to factor in the conformational heterogeneity within the unfolded-state.
### Figure 10.1

#### (a) Diagram of the system

- DA: donor-acceptor complex
- D*: excited donor
- A: ground state acceptor
- k_{exD}: excitation rate of donor
- k_{exA}: excitation rate of acceptor
- k_T: transfer rate
- k_D: diffusion rate
- R_0: equilibrium distance
- τ: lag time

#### (b) Kinetic Scheme

- DA → D*A
- k_{kD} → k_{kA}
- DA

#### (c) Transfer Efficiency

- Number of molecules: Static, Dynamic

#### (d) Lag Time Distribution

- g_{ij}(τ)

#### (e) Time Dependence

- Intensity I(t)
- Excitation pulse (t_0)
- Time, t (ns)

#### (f) Transfer Efficiency

- E(r) = \frac{k_{r}(r)}{k_{r}(r) + k_{D}}
- R_0

#### (g) Probability Distribution

- P(r)

#### (h) PMF

- G(r) = -k_B T \ln P(r)

#### (i) PMF in PMF

- Diffusion in PMF

---

**Equations:**

- \langle E \rangle = \int_{0}^{L} P(r) E(r) dr
- I(t) = \int_{0}^{L} P(r) e^{-dr/t} dr
- \frac{\partial p(r,t)}{\partial t} = D \frac{\partial^2 p(r,t)}{\partial r^2} - \frac{1}{P(r)} \frac{\partial P(r)}{\partial r} I + K_0(r) p(r,t)
The dynamics in proteins above the microsecond timescale usually involve persistent tertiary structure or cooperative interactions (41). In the absence of such folded or misfolded structures, the dynamics within unfolded or intrinsically disordered states are thus expected to be fast compared to the typical photon detection rates of $\sim 10^3 - 10^5$ s$^{-1}$ (137); this case results in a narrow, shot noise–dominated peak in the transfer efficiency histogram (56, 57). If the fluorophores reorient rapidly compared to the donor fluorescence lifetime, which is often the case for unfolded proteins and IDPs, as indicated by fluorescence anisotropy measurements (83, 113), $\chi^2$ can be assumed to average to 2/3, and the mean transfer efficiency of the unfolded state, $\langle E \rangle$, results as

$$\langle E \rangle = \int_{0}^{\ell} E(r)P(r) \, dr,$$

where $\ell$ is the distance of closest approach of the dyes, $\ell$ is the contour length of the polypeptide segment probed, and $P(r)$ is a suitably normalized probability density function (Figure 1). [If these requirements are not met, different types of averaging need to be used (139, 170).]

Based on these considerations, the shot noise–dominated peak corresponding to the unfolded state in a transfer efficiency histogram can be analyzed only in terms of a single parameter, its mean FRET efficiency, $\langle E \rangle$, and Equation 3 is thus understood unless additional information about the functional form of $P(r)$ is available. Different approaches have been chosen to obtain suitable distance distributions. The simplest approximation is the use of polymer–physical models to reflect the underlying chain statistics—for example, the Gaussian chain (67, 109, 138, 142), worm-like chain (109, 114, 139), excluded volume chain (98, 116), or a weighted Flory-Fisk distribution (70, 180) (Figure 2). Equation 3 can then be solved numerically for the single adjustable parameter in these models (e.g., for the mean-squared end-to-end distance in the case of a Gaussian chain). Provided that the segment length between the dyes is much greater than the persistence length of the chain, and that the mean of the distribution is near the Förster radius, these distributions usually provide very similar results in terms of average chain dimensions (109). Alternatively, molecular simulations (at various levels of coarse graining) have been employed to obtain information about the shape of $P(r)$ (11, 81, 100, 110, 116, 117, 139) and to assess deviations from simple polymer models (116, 139) (Figure 2).

Additional experimental information about $P(r)$ is available from the analysis of fluorescence lifetimes (67, 90, 147) (Figure 1): Because the dynamics of disordered states (for the chain lengths accessible in single-molecule FRET) are much slower than the fluorescence lifetimes of the dyes ($\sim 1$ to 4 ns), the distribution of distances results in a distribution of fluorescence intensity decay rates (60). In confocal experiments employing time-correlated single-photon counting, a subpopulation-specific analysis of fluorescence intensity decays can thus be used to test the validity of specific functional forms of $P(r)$, ideally by a global analysis of both donor and acceptor transfer efficiency histograms (56, 57). The distribution of distances results in a distribution of fluorescence intensity decay rates (60). In confocal experiments employing time-correlated single-photon counting, a subpopulation-specific analysis of fluorescence intensity decays can thus be used to test the validity of specific functional forms of $P(r)$, ideally by a global analysis of both donor and acceptor transfer efficiency histograms (56, 57).

### Figure 1

Experimental observables and data analysis in single-molecule Förster resonance energy transfer spectroscopy of unfolded-state dimensions and dynamics. (a) The signal from individual molecules labeled with donor (D) and acceptor (A) fluorophores freely diffusing through the confocal volume is detected and $\langle h(t) \rangle$ analyzed in terms of the distance-dependent transfer efficiency between donor and acceptor. (c) Transfer efficiency histograms illustrate the separation of subpopulations (here, folded state at high E, unfolded state at intermediate E, and donor-only population at E $\approx$ 0). (d) The combination with fluorescence lifetime analysis $\langle e \rangle$ provides additional information about distance distributions, especially via subpopulation-specific fluorescence intensity decays, $I(t)$, and (g) their analysis in terms of distance distributions based on polymer models or simulations (Figure 2). (b) Chain reconfiguration times are accessible via nanosecond fluorescence correlation spectroscopy (FCS) and a global analysis of donor-donor (DD), acceptor-acceptor (AA), and donor-acceptor (DA) correlation functions in terms of diffusion in a potential of mean force (PMF) (g) based on the distance probability density function, $P(r)$. $K_0(\ell)$ is the rate matrix that describes the distance-dependent kinetics of interconversion between the electronic states ($\ell$), and $p(\ell)$ is the vector of their populations.

---

**Shot noise:** the variation in count rates about fixed means due to the discrete nature of the signal. In single-molecule Förster resonance energy transfer (FRET), its contribution is often important because of the relatively small number of photons detected per unit time.

**Excluded volume:** describes the net two-body interaction between monomers in a chain, including steric repulsion and charge- or solvent-mediated interactions.

**Persistence length:** a measure of polymer stiffness, defined as the characteristic length along the chain over which orientational correlations between segments decay.
### Gaussian chain

\[ P(r) = 4\pi r^2 \left( \frac{3}{2\pi r^2} \right)^{3/2} \exp \left( -\frac{3}{2} \frac{r^2}{\langle r^2 \rangle} \right) \]

### Worm-like chain (WLC)

\[ P(r, l_p, l) = \frac{4\pi l l_p}{l \langle 1 - (r/l)^2 \rangle^{3/2}} \exp \left( -\frac{3l_c}{4l_p(1 - (r/l)^2)} \right) \]
\[ f(l_p, l) = \frac{1}{\pi^{3/2} e^{-\alpha^2}} \left( 1 + 3/\alpha + 15/(4\alpha^2) \right) \text{ with } \alpha = \frac{3}{4} \frac{l_c}{l_p} \]

### Self-avoiding walk (SAW)

\[ P(r) = \frac{a}{(r^2)^{1/2}} \left( \frac{r}{(r^2)^{1/2}} \right)^{2.5} \exp \left( -b \left( \frac{r}{(r^2)^{1/2}} \right)^{2.5} \right) \]

### Sanchez

\[ \frac{l_p}{R_g} = \int \frac{P(r|R_g)}{R_g} \frac{P(r, \epsilon, R_{gw}) dR_g}{R_g} \]
\[ P\left( r|R_g \right) = \frac{1}{\delta \cdot R_g} \left[ \frac{r}{\delta \cdot R_g} \right]^2 - \frac{9}{4} \left( \frac{r}{\delta \cdot R_g} \right)^3 + \frac{3}{16} \left( \frac{r}{\delta \cdot R_g} \right)^5 \]
\[ P\left( R_g |\epsilon, R_{gw} \right) = Z^{-1} R_g^6 \exp \left( \frac{-7 R_g^2}{2 R_{gw}^2} + \frac{1}{2} N c \phi - N \frac{1 - \phi}{\phi} \ln(1 - \phi) \right) \]

**Figure 2**

(a) End-to-end distance probability density functions, \( P(r) \), of polymer models commonly used in data analysis for unfolded proteins. Note that in all cases, a single model parameter is adjusted to fit the experimental data (\( \langle r^2 \rangle \) is the mean-squared end-to-end distance, \( \ell_p \) the persistence length, \( \epsilon \) the mean interaction energy between amino acids); all other parameters are known independently. \( R_g \) is the root-mean-squared radius of gyration of the \( \Theta \) state (see the section Chain Dimensions and Scaling Laws), and \( \phi \) is the volume fraction of the chain. \( l_c \) is the contour length of the chain segment probed; \( N \) is the number of segments; \( \delta \) is a correction factor (180); and \( Z, a, \) and \( b \) are normalization factors (116). (b, c) Comparison of \( P(r) \) from these models (lines) with simulation results (symbols) of the denatured-state end-to-end distance distributions of protein L at (b) 2.4 M and (c) 6 M guanidinium chloride (GdmCl) using a C\(^a\)-side-chain model and the molecular transfer model (116, 117). Panels b and c modified from Reference 116.

Intensity decays (67, 90, 147). A helpful diagnostic is a two-dimensional representation of the donor fluorescence lifetime versus transfer efficiency from count rates (Figure 1), in which the existence of a distance distribution resulting from rapidly interconverting conformations shows up as a deviation from the linear dependence of \( \tau_{DA} \) on \( E \) expected for a single fixed distance (59, 78, 147).

Residual secondary structure or the formation of long-range interactions can lead to deviations from the distance distributions based on simple models or simulations, even if global aspects such as...
length scaling are preserved (48, 116). Experimentally, these effects can be quantified by mapping different segments of the protein (by varying the positions of the fluorophores) and testing the consistency with the models employed (67, 70, 110, 116, 147) for example, in terms of the scaling of the measured average distance with the length of the segment probed (67, 147). Alternatively, Stokes radii can be obtained from fluorescence correlation spectroscopy (FCS) measurements, and although hydrodynamic properties can be difficult to relate to polymer models quantitatively, they provide complementary information about changes in chain dimensions and length scaling (70, 71, 97, 144). Finally, combining single-molecule FRET with complementary methods sensitive to the presence of structure, such as circular dichroism spectroscopy (68) or NMR (under conditions in which interference with the signal from folded or structured states can be excluded), will become increasingly important.

### Quantifying Chain Dynamics

Single-molecule fluorescence can be used to probe conformational dynamics over a wide range of timescales (137). A benefit of single-molecule experiments is the possibility of extracting information on dynamics, even from equilibrium measurements, without requiring a synchronization of the system by perturbation methods, as frequently employed in ensemble experiments. For proteins, equilibrium dynamics have been investigated based on correlation functions (21, 39, 108, 112, 113, 115, 143, 147, 152), the analysis of broadening and exchange between subpopulations in FRET efficiency histograms (25, 58, 68, 138), and from fluorescence trajectories of immobilized molecules (25, 26, 87, 88, 121, 124, 125). In this way, both the equilibrium distributions and the kinetics can often be obtained from the same measurement. Alternatively, nonequilibrium single-molecule dynamics can be investigated by microfluidic or manual mixing experiments (137).

The reconfiguration dynamics of unfolded proteins and IDPs can be probed with nanosecond fluorescence correlation spectroscopy (nsFCS) combined with FRET (55, 112, 113). Distance fluctuations between the donor and the acceptor fluorophores then result in fluctuations in fluorescence intensity on the timescale of interconversion between the different configurations of the polypeptide chain. For relating the measured decay time of the fluorescence correlation curves to the reconfiguration time of the chain, the relative motion of the two labeled positions can be represented as a diffusive process in a potential of mean force that corresponds to the equilibrium dye-to-dye distance distribution of the unfolded protein (**Figure 1**). In this simplified description, the dynamics of the system are determined (a) by the shape of the potential [i.e., the free energy as a function of the distance between the dyes, $G(r)$, which is obtained directly from the distance distribution function, $P(r)$, as described above] and (b) by an effective diffusion coefficient, $D$, that determines the relaxation time in the potential (55, 63, 112, 150, 167) (**Figure 1**). This relaxation time corresponds to the reconfiguration time, $\tau_r$, of the chain segment between the dyes.

In summary, the single-molecule approach described here can provide two essential quantities that characterize the global properties of unfolded or intrinsically disordered proteins: intramolecular distance distributions (from transfer efficiency and fluorescence lifetime measurements) and reconfiguration dynamics (from fluorescence correlation analysis). These are key quantities in the statistical mechanics of polymers (20, 30, 38, 50) and thus make the results amenable to the application (and even the testing) of concepts in polymer physics, with the goal of providing a coherent and quantitative framework for the behavior of unfolded proteins and IDPs.
POLYMER PROPERTIES OF UNFOLDED AND DISORDERED PROTEINS

Denaturants and the Collapse Transition

Polymers respond to changes in solvent quality by a change in chain dimensions: In poor solvent, where intrachain interactions dominate over chain-solvent interactions, they collapse into compact globules; in good solvent, where chain-solvent interactions are favorable, they populate rather open, expanded conformations (20, 49). In water, many proteins fold into well-defined, compact structures and/or tend to aggregate; in solutions of denaturants such as urea or guanidinium chloride (GdmCl), however, proteins and peptides are highly soluble and (in the absence of disulfides) form expanded unfolded states (80, 127). Correspondingly, the chain dimensions of folded globular proteins, quantified, e.g., in terms of the radius of gyration, \( R_g \), exhibit a dependence on the number of peptide segments in the chain, \( N \), that is in good agreement with the scaling behavior expected for compact globules, \( R_g \propto N^{1/3} \) (36, 159, 168). For unfolded proteins in high concentrations of denaturant, measurements of intrinsic viscosity, NMR, and small-angle scattering have shown good agreement with the length scaling of chain dimensions with \( N^{3/5} \) (84, 154, 155, 168), as expected for a polymer in good solvent.

From a polymer physics perspective, even in the absence of a transition to a folded structure or populated folding intermediates, denatured states are expected to exhibit a compaction akin to heteropolymer collapse when the solvent quality (here modulated by the denaturant concentration) is reduced (2, 20, 27, 71, 117). This behavior has been demonstrated particularly clearly by the molecular transfer model (117) (Figure 3b), which combines simulations in the absence of denaturants with experimentally determined transfer free energies of side chain and backbone moieties from water to denaturant solution (153). However, detecting this behavior in ensemble experiments had been challenging, as the corresponding signal change is hard to separate from the folding transition. The separation of folded and unfolded subpopulations in single-molecule FRET experiments (33) led to the identification of this unfolded-state compaction in terms of a continuous shift in transfer efficiency (138, 142) (Figure 3a); it has since been observed for many proteins (65, 137) and is in agreement with a range of other experiments, theory, and simulations (71, 94, 107, 117, 155, 156, 158, 174, 181). However, the detectability of unfolded-state compaction in SAXS experiments has remained controversial (175).

Particularly interesting is the analysis of denaturant-dependent unfolded-state compaction in terms of the theory of coil-globule transitions (35, 61, 131), which allows chain compaction or expansion to be related to changes in effective intrachain interaction energies between the residues. Haran and coworkers (142, 180) used the theory of Sanchez (131), which provides an expression for the probability density function of \( R_g \) in the form of a Boltzmann-weighted Flory-Fisk distribution (Figure 2). Ziv & Haran (180) analyzed a large set of results in which the mean transfer efficiency of unfolded proteins had been obtained over a broad range of denaturant concentrations (Figure 3c,d). By determining the mean interaction energy between amino acids from \( \langle E \rangle \) and computing the free energy of the chain based on the Sanchez theory, they found that the difference between the free energy of unfolded proteins and their most collapsed states depends approximately linearly on the concentration of GdmCl and that the change of this free energy with denaturant concentration is close to the equilibrium \( m \) value, the derivative of the free energy difference between the folded and unfolded states with respect to the denaturant concentration. Ziv & Haran concluded that a major part of the free energy change for protein folding originates from the free energy change of collapse, thus providing a link between the polymeric properties of the unfolded state and protein folding mechanisms (180).
Charge Interactions and Polyampholyte Theory

Proteins are heteropolymers, and their effective intrachain interaction energy, as reflected by the mean interaction energy between amino acids as expressed in the Sanchez theory (Figure 2), comprises different contributions, including hydrogen bonding and hydrophobic and electrostatic interactions. For IDPs, charge interactions are expected to be particularly pronounced, as their
amino acid composition is strongly biased toward low hydrophobicity and high charge content (97, 161). The conformational properties of proteins, especially their degree of expansion, as a function of these contributions can be classified in terms of a state diagram (29, 97, 164) (Figure 4a). One way of treating the underlying interactions of charges within the chain quantitatively is through polyampholyte theory (37), which encompasses both repulsive interactions between charges of the
same sign and attractive interactions between charges of opposite sign. The approach of Higgs & Joanny (66), for example, considers a chain in which monomers carry a positive charge with probability $f$ and a negative charge with probability $g$. The monomers have an excluded volume $v_b^3$ owing to steric repulsion (with a segment length $b = 0.38$ nm, the $C^\alpha-C^\alpha$ distance in a polypeptide) and interact via screened Coulomb interactions between each pair of monomers, with distances assumed to follow Gaussian chain statistics. The effect of the electrostatic interactions on chain dimensions can then be treated as a contribution to an effective excluded volume, $v^*b^3$. Introducing the expansion factor $\alpha$ as the ratio of an effective segment length $b_1$ and the real segment length $b$, one can then describe the chain dimensions using the expressions

$$\alpha^2 - \alpha = \frac{4}{3} \left( \frac{3}{2\pi} \right)^{1/2} N^{1/2} \nu^*, \quad \text{where} \quad \alpha = \frac{b_1}{b},$$

and

$$v^*b^3 = vbb^3 + \frac{4\pi l_\delta(f - g)^2}{\kappa} - \frac{\pi l_\delta(f + g)^2}{\kappa}.$$

The second term of the right-hand side in Equation 5 accounts for repulsive interactions due to the net charge of the polypeptide, resulting in an increase of $v^*b^3$, as in related forms of polyelectrolyte theory (62). The third term leads to a reduction of $v^*b^3$ through attractive interactions between charges of opposite sign. $l_\delta$ is the Bjerrum length and $1/\kappa$ the Debye length.

Müller-Spåth et al. (109) applied this theory to a series of unfolded and intrinsically disordered proteins for which the importance of the net charge was qualitatively obvious from the different response of chain dimensions with urea, which is uncharged, and with GdmCl, which is a salt (69) (Figure 4b): With increasing urea concentration, a monotonic expansion was observed for all chains, but the screening of charged residues in the IDPs by GdmCl caused a pronounced collapse of charged IDPs between 0 and 0.5 M (Figure 4b), as expected for a polyelectrolyte (62, 118). Remarkably, however, unfolded cold shock protein (Csp), which contains approximately equal numbers of acidic and basic amino acids, exhibited signs of polyampholyte behavior: The addition of salt led to an expansion rather than a collapse, indicating that attraction between charge residues of opposite sign causes compaction. The response of the entire set of proteins was described well by the theory of Higgs and Joanny, combined with a simple binding model (96) accounting for denaturant-induced chain expansion (109). The result demonstrated that charge interactions can dominate the dimensions of unfolded proteins and IDPs, in accord with molecular simulations (97), and that a mean-field theory can account for these observations. However, charge patterning within the sequence (29) or the formation of persistent local or long-range structure (44, 93) can lead to deviations from this relatively simple behavior (Figure 4c). For instance, segregation of charges along the sequence can lead to structure formation, and simulations or more advanced theoretical approaches are required for a quantitative treatment of these effects (29, 132).

**Chain Dimensions and Scaling Laws**

Denaturants and charge interactions are only two of the factors that have a pronounced influence on the dimensions of denatured proteins and IDPs. Given the quasi-continuous spectrum of their amino acid compositions (161) and the broad range of relevant solution conditions, a quantitative way of classifying the degree of expansion or collapse of unfolded and intrinsically disordered proteins is essential (Figure 4a). That some denatured proteins are neither as expanded as fully unfolded states nor as compact as folded proteins was already realized in the 1980s. Both equilibrium and kinetic intermediates in protein folding with such “molten globule” properties were observed (123), and the scaling of chain dimensions with the length of the polypeptide was found

**Polyampholyte theory:** describes the effect of charge interactions on the conformations of a polymer, including both repulsive interactions between charges of the same sign and attractive interactions between charges of opposite sign

**Polyelectrolyte:** a macromolecule in which a substantial fraction of the monomers contain ionizable or ionic groups

**Polyampholyte:** a polyelectrolyte that contains both positively and negatively charged groups
**Scaling laws:** describe the scale invariance observed for many natural phenomena and are frequently used in polymer physics, for example, to quantify the scaling of chain dimensions with chain length.

**Θ conditions:** conditions under which attractive interactions within the chain counterbalance the steric repulsion between monomers, and the polymer follows the statistics of an ideal chain to good approximation.

One possibility is to relate the denaturant-dependent collapse and other effects on chain compaction quantitatively to the effective interaction energies between the residues via the Sanchez theory of coil-globule transitions described above. Doing so requires knowledge of the dimensions of the chain at its Θ point, where chain-chain and chain-solvent interactions of a real chain balance in a way that its length scaling approaches that of an ideal chain (i.e., $R_g/\Theta \propto N^{1/3}$). However, systematic variation in the number of monomers over a large range—the classic way of determining length scaling in polymers—is difficult for proteins. Hofmann et al. (70) thus made use of the finding that previously determined values of the prefactor, $R_g/\Theta$, in scaling laws of the type $R_g = R_{\Theta} N^{\nu}$ vary only from $\sim 0.2$ nm to $\sim 0.3$ nm, even when comparing polypeptides in high concentrations of denaturants and folded globular proteins (48, 84, 168, 176), and can be interpolated in this range (70). With this constraint, the resulting scaling exponents, $\nu$, can be quantified for a broad range of proteins (Figure 4d) and depend sensitively on their sequence composition, net charge, and denaturant concentration. This approach also allows the Θ conditions to be identified based on the chain dimensions. Interestingly, protein sequences with a moderate net charge and average hydrophobicity were found to be close to Θ conditions in water (70). It has been suggested that folding from the Θ state provides an ideal evolutionary compromise by favoring intrachain interactions while allowing for rapid sampling of different conformations (17).

Length scaling exponents of unstructured peptides close to 1/3, as found for polyglutamine (27), are in line with the high aggregation tendency of these pathogenic sequences. Based on single-molecule FRET experiments of an FG-repeat nucleoporin, a connection between chain collapse and the capacity to form hydrogels has been proposed (103). In summary, length scaling exponents favoring intrachain interactions while allowing for rapid sampling of different conformations (17).

**Temperature-Induced Compaction and Solvation Free Energies**

The dimensions of unfolded proteins and IDPs show a response to temperature that may be somewhat counterintuitive. Typical homopolymers in organic solvent, for example, expand monotonically with increasing temperature, as expected if thermal energy increases compared with attractive intrachain interactions (131). For proteins, however, laser temperature-jump experiments had indicated a compaction of the acid-denatured protein BBL with increasing temperature (129), and NMR measurements of the C-terminal domain of protein L9 in the cold-denatured state had shown increasing radii of hydration as temperature was reduced (92). The separation of folded and unfolded subpopulations in single-molecule FRET experiments enabled an investigation of this issue over a wider range of temperatures. For Csp, a continuous collapse of the unfolded state between $\sim 280$ K and $340$ K was observed (114). Similarly, denatured frataxin collapsed over a broad range of temperatures, spanning both the cold- and the heat-denatured state (4).

Although this temperature-induced compaction is suggestive of hydrophobic interactions as a driving force (92, 129), highly charged IDPs exhibit the most pronounced collapse (114, 172), indicating the importance of additional contributions related to changes in solvation free energy as a function of temperature. A critical role of the solvent contribution is supported by molecular simulations of unfolded proteins with different water models (12, 114). Interestingly, the compaction observed for five different unfolded or intrinsically disordered proteins could be...
Figure 5
Effects of crowding and cellular conditions on intrinsically disordered proteins. (a) Crowding effect on the conformation of a self-avoiding-walk chain as a function of crowder volume fraction, $\phi$, and the ratio of chain dimensions and the crowder size, $\lambda$, for solid spherical crowders based on simulations (79). An illustration of the chain and the crowding particles for $\lambda = 1.9$ and $\phi = 0.3$ is shown on top. (b, top) Graphical representation of scaled-particle theory (SPT), Gaussian cloud model (gcSPT), Flory-Huggins theory (FH) in the short-chain regime, and renormalized Flory-Huggins theory (renormalized FH) in the long-chain regime. (b, bottom) Radius of gyration of ProTαC as a function of the degree of polymerization of PEG at $\phi = 0.15$. Fits according to the different theories are shown as black (SPT), gray (gcSPT), cyan (FH), and blue (renormalized FH) lines. Solid lines indicate the regime for which the respective theories were derived; otherwise, dashed lines are used. (c) Single-molecule Förster resonance energy transfer (FRET) measurements in live cells after microinjection of labeled unfolded molecules: (top) Schematic illustration, (middle) single-cell FRET efficiency histograms of prothymosin α (ProTα) in the cytosol and nucleus of HeLa cells, and (bottom) nanosecond fluorescence correlation spectroscopy (nsFCS) measurements of chain dynamics in live cells (83). Panel a modified from Reference 79, panel b modified from Reference 148, and panel c modified from Reference 83.

Crowding and Excluded-Volume Screening
Ultimately, we want to understand the behavior of unfolded proteins and IDPs in a cellular environment. An important difference compared to typical experiments in vitro is the high concentration of cellular components, resulting in molecular crowding effects (177). In view of their lack of persistent structure, IDPs and unfolded proteins are expected to be particularly sensitive to the effects of crowding (Figure 5a). Indeed, experiments indicate that some unfolded proteins or IDPs gain structure upon crowding (32), while others do not (99, 151) but may change their dimensions (72, 77, 101, 148). The theoretical concept most widely used for rationalizing crowding

Molecular crowding: describes the effects of high concentrations of inert cosolutes on molecular interactions and conformations

reproduced semiquantitatively in implicit solvent simulations by taking into account the solvation free energies of the constituent amino acids, with the solvation properties of the most hydrophilic residues playing a large part in determining the collapse (172). Temperature-dependent studies can thus provide an important link between unfolded-state dimensions and the solvation properties of the constituent amino acids, similar in spirit to the molecular transfer model (117).
Chain Dynamics of Unfolded and Intrinsically Disordered Proteins

Conformational dynamics in proteins can occur over an enormous span of timescales owing to the broad range of enthalpic and entropic contributions involved. The slowest conformational change for an individual structured protein will usually be its global folding and unfolding, which can take place on timescales from microseconds to hours and beyond. Within unfolded or intrinsically disordered states, however, cooperative interactions linked to the formation of specific structures are much less prevalent; correspondingly, conformational dynamics are expected to be much faster.

Unexpectedly slow dynamics up to the seconds timescale have been reported both for chemically unfolded ribonuclease H (87) and for some proteins previously classified as intrinsically disordered (24), but the structural origins (including the possible role of peptidyl-prolyl cis-trans isomerization) have remained unresolved. For α-synuclein, slow dynamics are induced by the interaction with membranes (47).

In the absence of persistent intra- or intermolecular interactions, an unfolded polypeptide chain is expected to approach dynamics that are limited by monomer diffusion through the solvent and described by simple polymer models, such as the Rouse or Zimm theories (38). In the Rouse theory, chain dynamics are represented by harmonic modes, arising from the motion of beads representing individual chain segments connected by ideal springs, and whose dynamics are thus controlled only by the neighboring beads and the viscous drag from the solvent. For several proteins unfolded in high concentrations of denaturant, good agreement with these models has been observed, corresponding to reconfiguration times, $\tau_r$, of several tens of nanoseconds for 50- to 70-residue segments (147). Note that according to Onsager’s regression hypothesis, this time also corresponds to the collapse time of the chain (112), in agreement with laser temperature-jump experiments (129), and is thus closely linked to the pre-exponential factor in protein folding
Internal friction:
a dissipative force within
a medium (in contrast
to hydrodynamic
friction with the
solvent), here within
an unfolded protein,
that slows down its
dynamics (40, 112, 129, 173). In terms of an effective end-to-end diffusion coefficient (150), these
dynamics are in a similar range as those of unstructured peptides (39, 63, 89, 107).

However, with decreasing denaturant concentration, $\tau_i$ (corrected for solution viscosity) typ-
ically increases concomitant with chain compaction (14, 112, 147), in contrast to the behavior
expected from the Rouse or Zimm model, where chain compaction leads to a decrease in $\tau_r$ (38).
This discrepancy is a clear indication for the presence of internal friction (112). Defined operation-
ally as a deviation from proportionality between relaxation time and solvent viscosity, internal
friction had previously been identified in folded proteins (3) and in the context of protein folding
reactions (18, 64). Its molecular origin was mainly assigned to the large degree of desolvation in
the very compact folded or transition states and the resulting decoupling from the solvent. For
unfolded proteins, however, where the expansion of the chain relative to the folded state entails
pronounced solvation even in the absence of denaturant (67, 70), the origin of internal friction is
less obvious.

Internal Friction

There is a long history of theoretical concepts in polymer dynamics that address the question of
internal friction (or internal viscosity). Early ideas go back to Kuhn & Kuhn (85), Cerf (19), and de
Gennes (30). A particularly simple class of models that allow internal friction to be incorporated
rigorously (23) are those of Rouse (128) and Zimm (38, 178). Internal friction can be included
in terms of an internal friction coefficient that impedes the relative motion of two beads (7, 82)
(Figure 6a). Notably, this additional term does not change the eigenmodes of the system, but it
increases all relaxation times by the same internal friction time, $\tau_i$, yielding a spectrum of relaxation
times,

$$\tau^{(n)} = \tau_{Rouse}/n^2 + \tau_i, \quad n = 1, 2, \ldots,$$

where $\tau_{Rouse}$ is the Rouse time, corresponding to the slowest relaxation mode of the Rouse chain
in the absence of internal friction, and $n$ is the mode number. A simple common assumption is
that only the dynamics corresponding to $\tau_{Rouse}/n^2$ depend on solvent viscosity, $\eta$, whereas $\tau_i$ does
not (30). Because the overall solvent-dependent relaxation time, $\tau_{fr}$, with contributions from all
relaxation modes, is the same as $\tau_{Rouse}$ to within a numerical factor (95), we arrive at the relation

$$\tau_{fr}(\eta) = \tau_{fr}(\eta_0)\eta/\eta_0 + \tau_i,$$

where $\eta_0$ is the viscosity of water. In other words, if the chain reconfiguration time is measured
as a function of solution viscosity and extrapolated linearly to zero viscosity, the internal friction
time would result as the intercept. The same behavior is found for the Zimm model when inter-
nal friction is included analogously, only with a mode dependence of $n^{-2/3}$ compared to $n^{-2}$ in
Equation 6 (23).

The solvent viscosity dependence predicted by Equation 7 is in good agreement with experi-
mental observations for unfolded Csp (147), with increasing values of $\tau_i$ as the denaturant
concentration is decreased (Figure 6b). This result was confirmed by an orthogonal approach not
requiring changes in viscosity: Because the chain exhibits a spectrum of relaxation modes (Equa-
tion 6), the relative contribution of internal friction to the reconfiguration dynamics will depend
on the length of the segments probed (82, 95). By placing the FRET dyes in different positions
within the chain, one can determine $\tau_i$ independently. Very similar results are obtained (147) using
the Zimm model with internal friction (23). For chemically unfolded spectrin domains, significant
internal friction was present even at the highest GdmCl concentrations (14), indicating a possible
role for residual local interactions and secondary structure formation. Notably, spectrin domains
Figure 6

Internal friction in unfolded proteins and intrinsically disordered proteins (IDPs). (a) Schematic of a Rouse chain (represented as beads coupled by harmonic springs) with internal friction (indicated as dashpots) and terminal dyes. (b) Solvent viscosity ($\eta$) dependences of chain reconfiguration times, $\tau_{r}$, of terminally labeled cold shock protein at different guanidinium chloride concentrations (indicated in each panel) and fits with the Rouse model with internal friction. The values of the internal friction time, $\tau_{i}$, correspond to the intercepts (red arrows). (c) Results analogous to panel b from atomistic simulations (42). (d) The dependence of $\tau_{i}$ on chain expansion of different unfolded proteins and IDPs, as indicated by their apparent persistence length, $l_{p}$, shows a trend of more internal friction for more compact chains [prothymosin $\alpha$ (blue), integrase (yellow), Csp (red), spectrin R17 (dark green), spectrin R15 (light green)] (147). (e) End-to-end contact formation time, $\tau_{esp}^{(4)}$, of a (GlySer)$_{4}$ peptide from atomistic simulations where viscosity is varied by rescaling the solvent mass (140), with linear (solid line) and power-law (dashed line) fits, and a fit based on a Rouse model for contact formation including internal friction (dotted line). (f) The low viscosity dependence of the isomerization rate for a four-bead butane-like molecule (inset) indicates that solvent memory effects can contribute to the signature of internal friction observed experimentally (31). Panel a modified from Reference 82; panel b modified from Reference 147; panel c modified from Reference 42, copyright 2014, American Chemical Society; panel e modified from Reference 140, copyright 2012, American Chemical Society; and panel f modified with permission from Macmillan Publishers Ltd: Nature Communications, copyright 2014, Reference 31.

are $\alpha$-helical proteins, whereas Csp is an all-\beta protein, supporting suggestions based on simulations that interactions such as hydrogen bonds (140) local in sequence contribute more strongly to internal friction in unfolded proteins than the highly nonlocal interactions characteristic of $\beta$ structure (31).

What is the role of internal friction in IDPs? Both the N-terminal domain of HIV integrase and the highly negatively charged prothymosin $\alpha$ (ProT$\alpha$) exhibit much less internal friction than unfolded Csp or spectrin in the absence of denaturant (147). A plot of $\tau_{i}$ versus the dimensions of the chains in terms of the radius of gyration shows an overall trend corresponding to an increase in $\tau_{i}$ with chain compaction (Figure 6d), suggestive of the general importance of intrachain interactions for internal friction. Simulations have started providing important insights as to the molecular origin of internal friction. Echeverria et al. (42) simulated unfolded Csp at different GdmCl concentrations, found remarkable agreement with the experimentally determined dynamics (Figure 6c), and concluded from their analysis that dihedral angle transitions provide the dominant mechanism of internal friction.
connection to intrachain interactions such as hydrogen bonding, which has been suggested to cause internal friction based on atomistic simulations of peptides (140) (Figure 6e). Another potentially very important contribution is the low sensitivity of dihedral angle isomerization to solvent viscosity, as predicted theoretically by Portman et al. (122) and recently identified to affect protein folding dynamics by de Sancho et al. (31) in simulations (Figure 6f).

CURRENT DEVELOPMENTS AND FUTURE CHALLENGES

Polymer concepts have proven very valuable for a physical understanding and classification of the structural and dynamic behavior of unfolded and intrinsically disordered proteins. This is the case especially in the context of the global properties typically probed with single-molecule FRET. Obviously, the applicability of simple models needs to be evaluated for each case, and in the presence of persistent secondary or tertiary structure, the role of specific interactions and protein folding may have to be included in a physical description. For many questions, a closer combination with experimental methods complementary to single-molecule spectroscopy will thus be important, especially circular dichroism, scattering techniques (9, 53), and NMR (75). Although often limited to conditions in which the unfolded or disordered state is populated exclusively, these methods can provide essential additional information, such as the local structural and dynamic properties available from NMR. This type of integrated approach will lead to a more comprehensive view of the behavior of unfolded proteins and IDPs (162). In regard to single-molecule spectroscopy, challenges include developing a more complete use of the broad span of timescales accessible from correlation analysis (136, 146), taking full advantage of three-color FRET experiments (52, 102), and the extension to intracellular measurements (1, 83).

At the same time, simulations will become increasingly important for linking experimental observations and chain statistics to molecular detail. The traditional focus on folded proteins for the parameterization of the energy functions for molecular dynamics simulations has typically resulted in unfolded states that were much more compact than observed experimentally (13, 120). The increasing interest in IDPs, however, has triggered developments to remedy this deficiency, in particular the optimization of suitable implicit (166, 172) and explicit water models (12, 13, 111, 120), and also the explicit representation of the fluorophores (10, 11, 100, 135). The more routine availability of such simulations ideally complements the further development of analytical theory (6, 86, 132) and will enable us to obtain increasingly accurate representations of unfolded proteins and IDPs even though their heterogeneous conformational ensembles are underdetermined by experimental observables alone (75, 110, 165).

This improved quantitative understanding paves the way for addressing questions of increasing complexity, such as the interactions of IDPs with their biological targets, be they other proteins, nucleic acids, lipids, small molecules, or metal ions (171). For instance, many IDPs are known to remain partially, largely, or possibly even completely unstructured in the bound state (157) and thus retain much of their polymer dynamics. Closely connected are the processes of liquid-liquid phase separation and coacervation (16), which have recently seen a revival in the context of non-membrane bound compartments in cells (74). A key challenge will be to bridge the gap between the mesoscopic picture emerging from the concepts of phase separation (74) and the underlying microscopic/molecular mechanisms (91), both of them deeply rooted in polymer physics. Methods such as single-molecule spectroscopy and NMR are ideally suited for characterizing the resulting heterogeneous conformational ensembles and understanding the mechanisms of complex formation. In summary, a large array of problems related to the conformations and dynamics of unfolded and intrinsically disordered proteins is waiting to be addressed, and concepts from polymer and soft matter physics will be essential for progress.
DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Alessandro Borgia for discussions and for providing raw data for figures, and all members of our research group for discussions over the years. Our work has been supported by the Swiss National Science Foundation, the European Research Council, and the Human Frontier Science Program.

LITERATURE CITED

29. Das RK, Pappu RV. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. PNAS 110:13392–97


